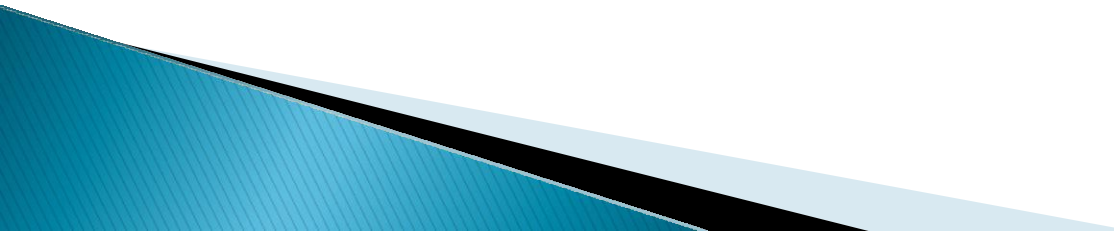


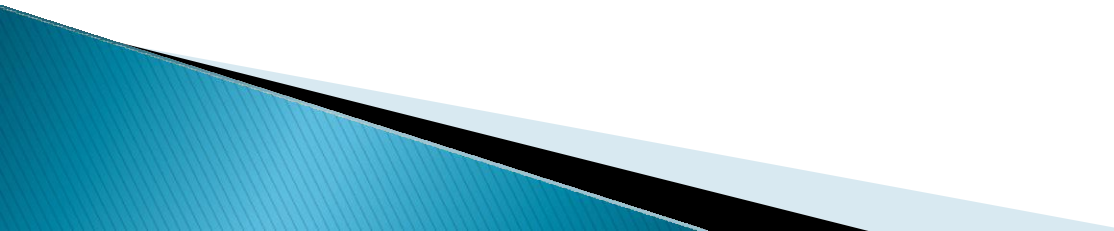
Concevez une application au service de la santé publique

Par : BAHRI
Abdelghani

SOMMAIRE

- Contexte et objectifs
 - Une alimentation saine ?
 - L'analyse univariée/ multivariée
 - Application
 - Réduction dimensionnelle
 - Conclusion
- 

Contexte et objectifs

- Création d'un générateur de recettes saines
 - Enjeu primordial pour les industries agroalimentaires
 - Définir ce que veut dire l'alimentation saine
 - Analyser les données nutritionnelles des produits.
- 

Une alimentation saine ?

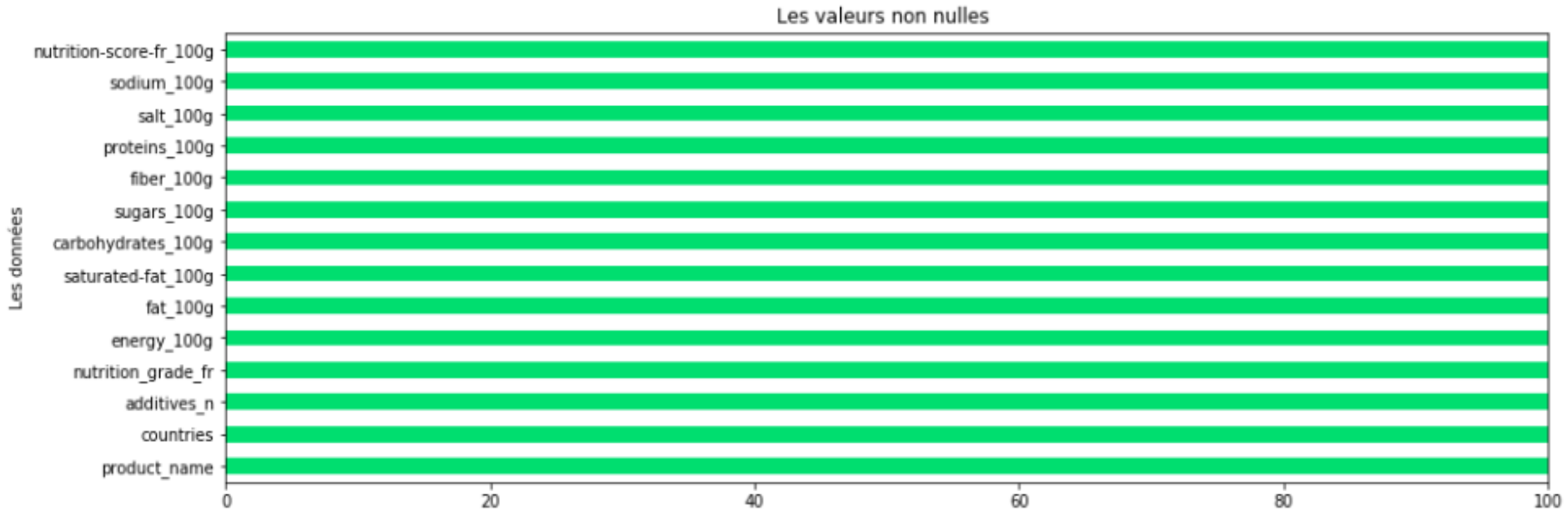
□  des fibres, Protéines, Vitamines,
Calcium, Fer

□  de Calories, Cholestérol,
Gras, Sucres, Sodium

Nettoyage des données

- Supprimer les données avec plus de 60% de valeur NaN en fonction des méthodes statistiques utilisées
- Supprimer les champs inutiles qu'on a pas besoin pour cette analyse nutritionnelle
- Traiter les données NaN
 - « inconnu » : pays, nom de produit
 - 0 : le score nutritionnelle
 - Supprimer le reste des NaN

Visualiser la consistance des données non nulles

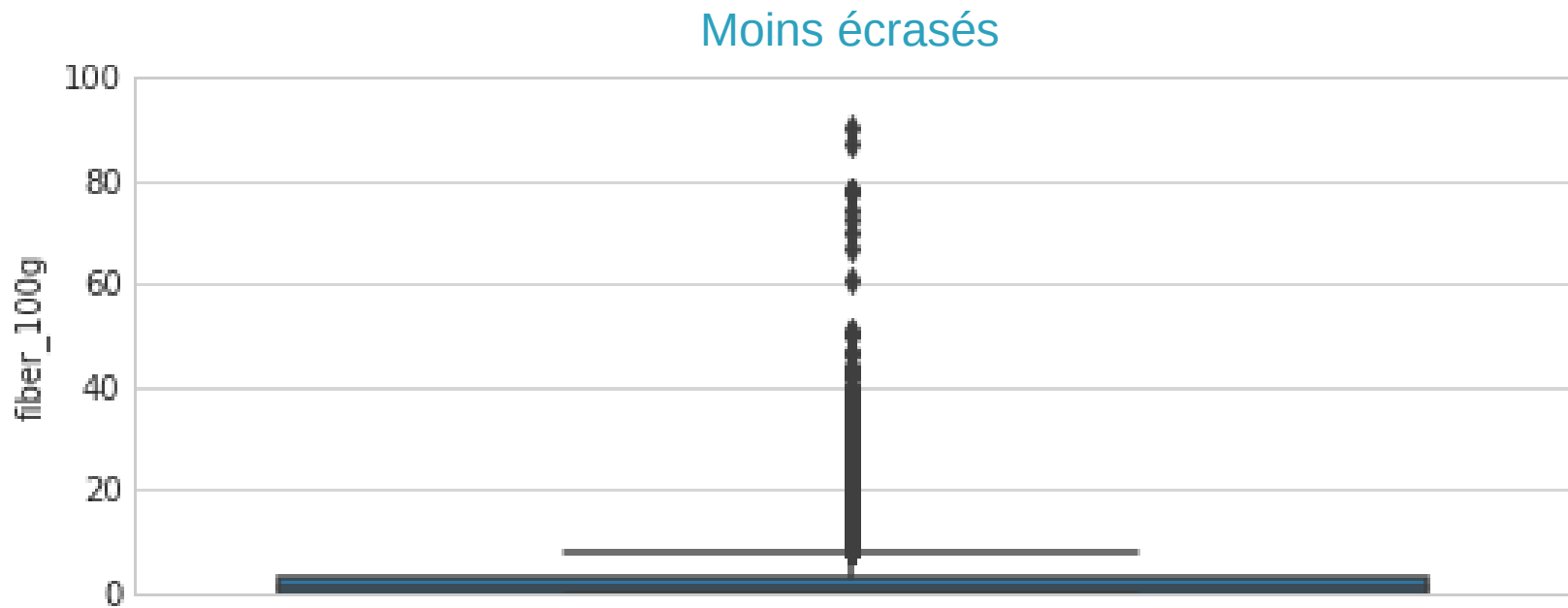


L'analyse des données nutritionnelles

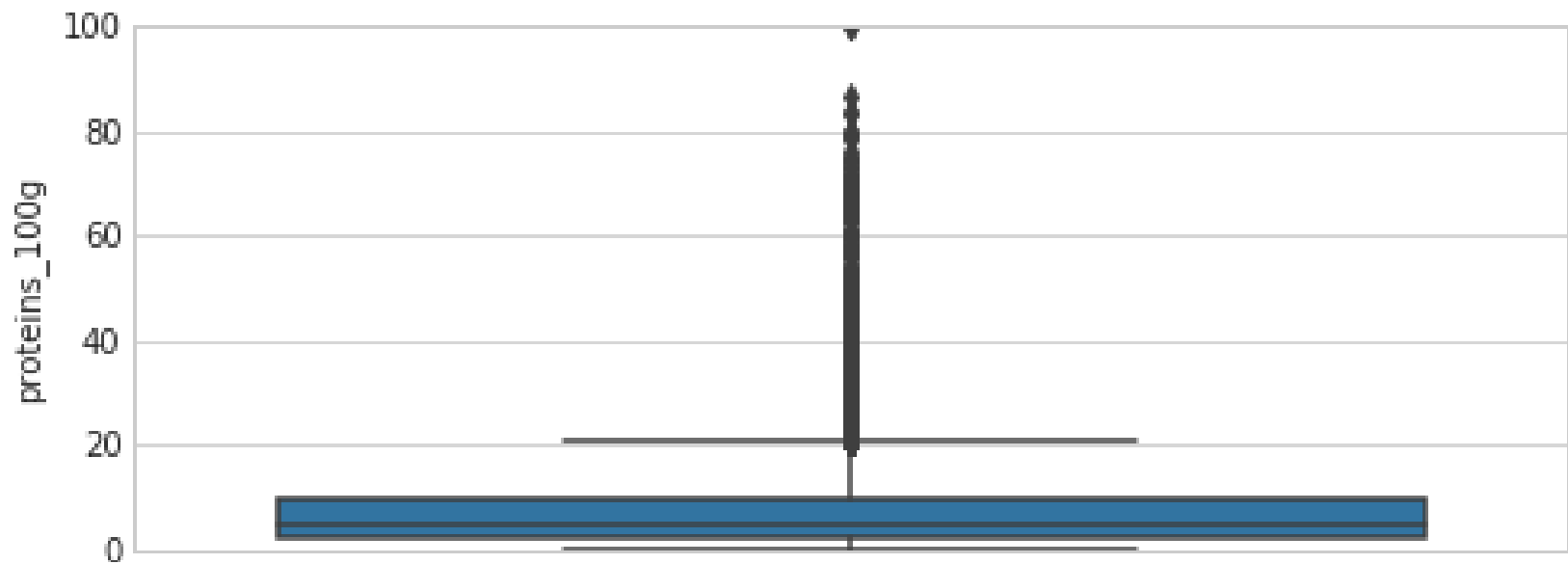
- L'analyse uni-variée
- L'analyse multi-variée

L'analyse univariée

Fibre vs Protéine

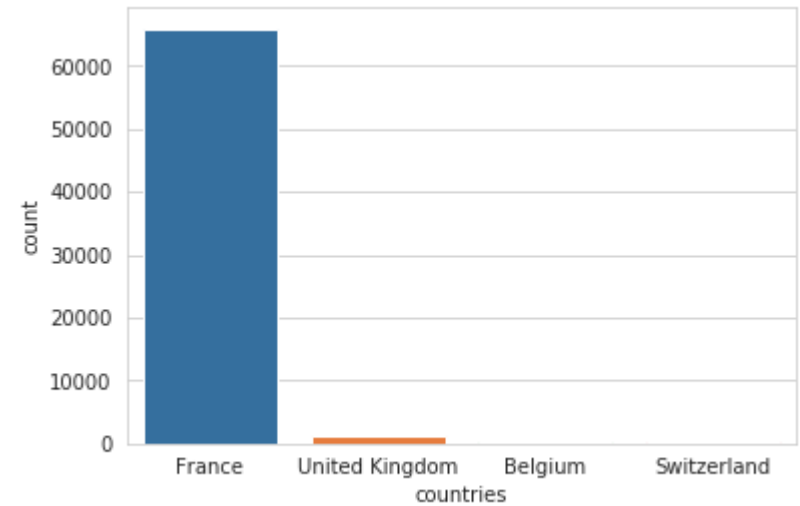
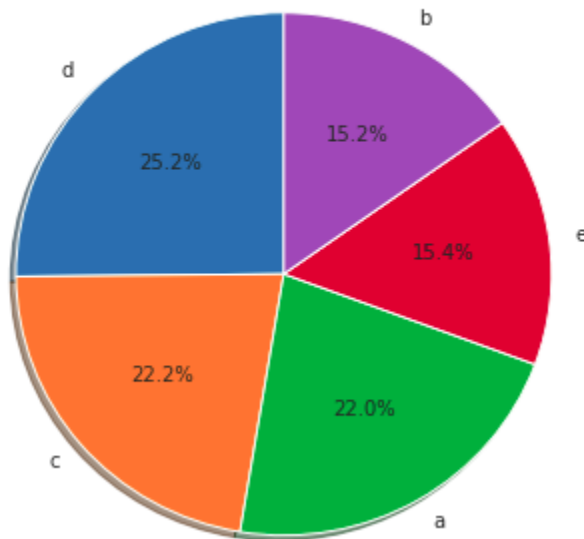


Un peu bancal



L'analyse des données nutritionnelles

Répartition des produits selon leurs classement de nutrition

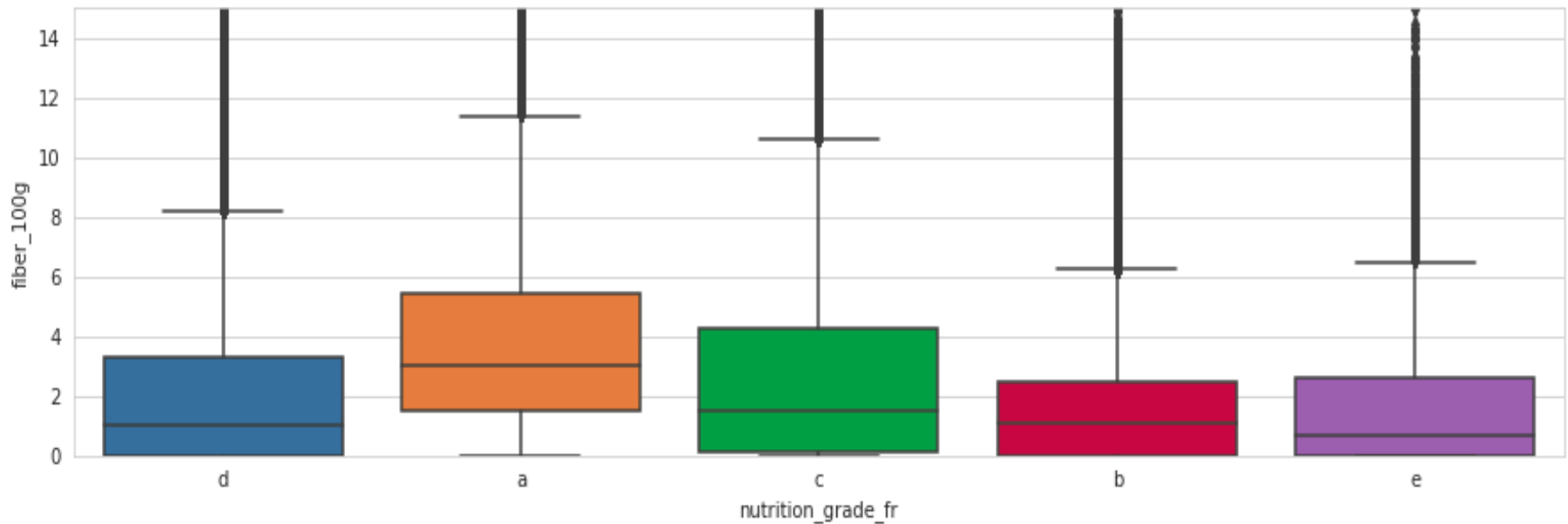


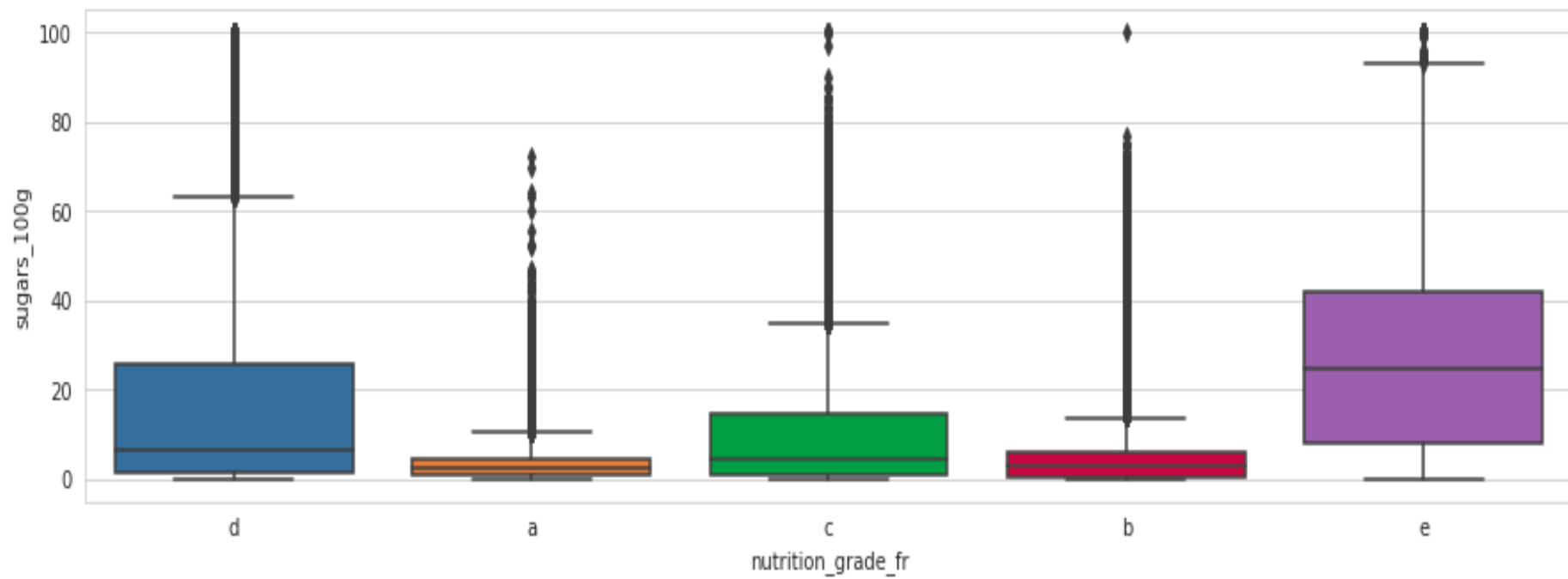
La France "le pays qui produit le plus"

Le score nutritionnel

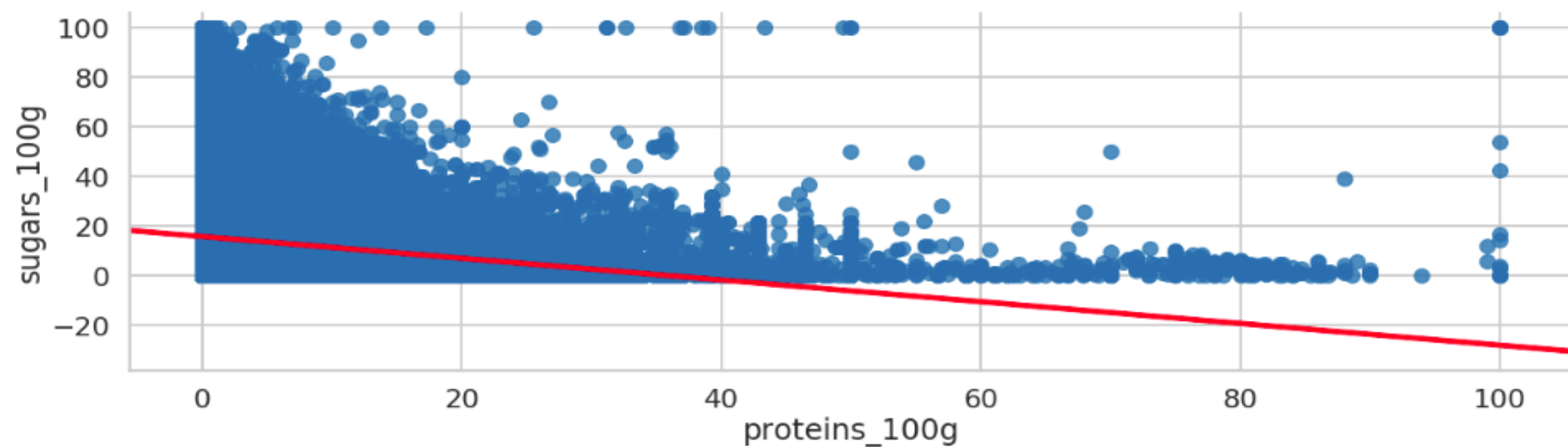
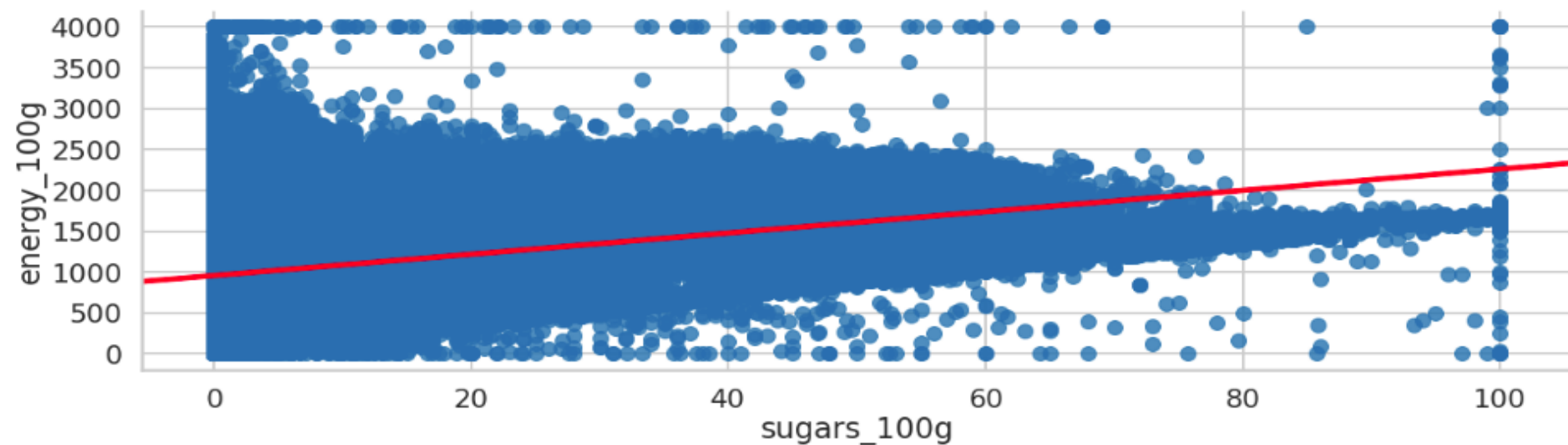
- Le Nutri-Score, c'est un système basé sur une échelle de 5 couleurs, associé à des lettres allant de A à E.

L'analyse multivariée

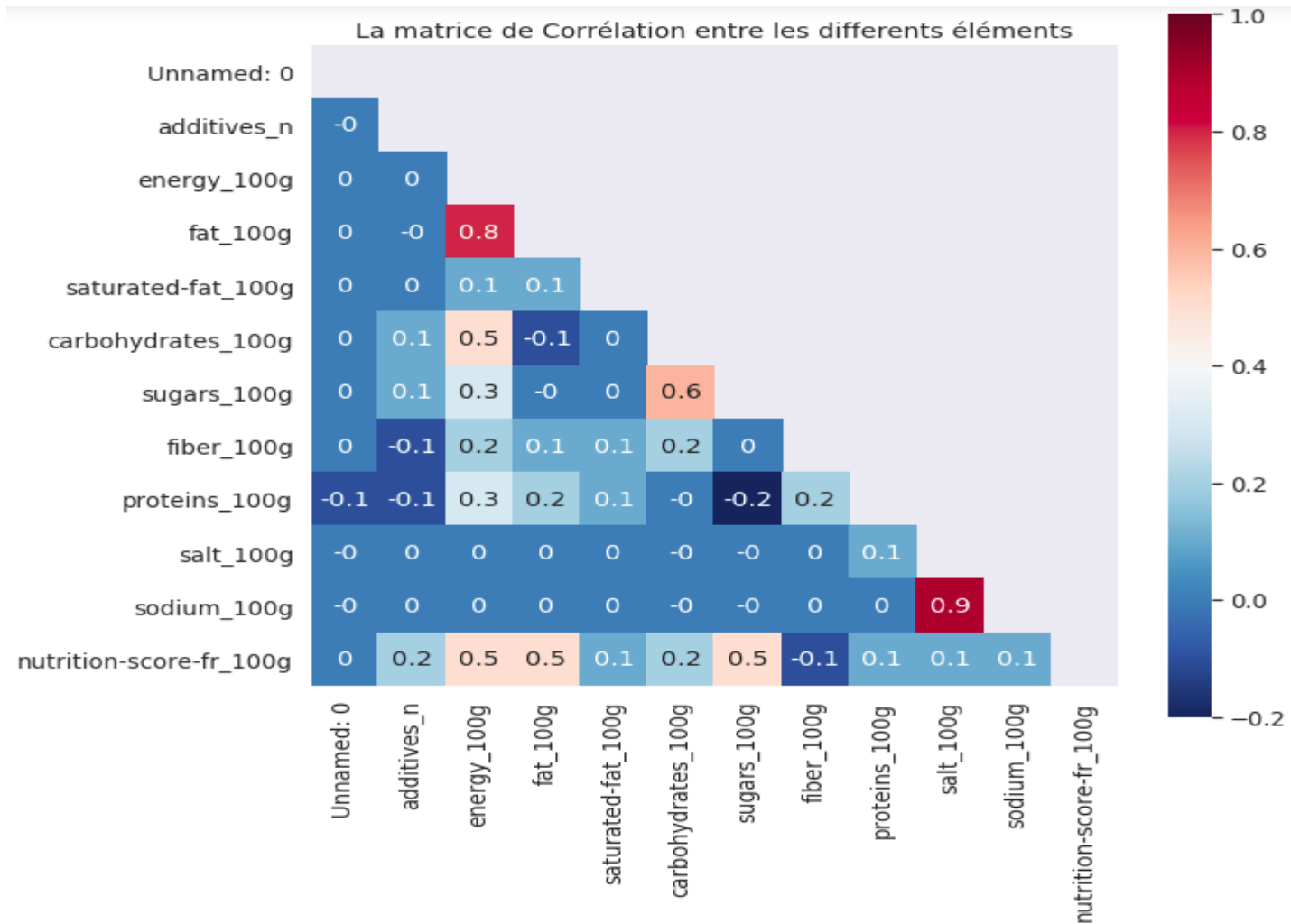




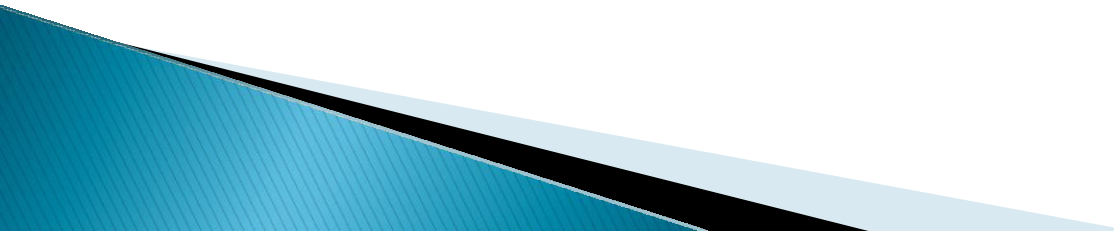
Représentation des données selon deux dimensions



Matrice de Corrélation



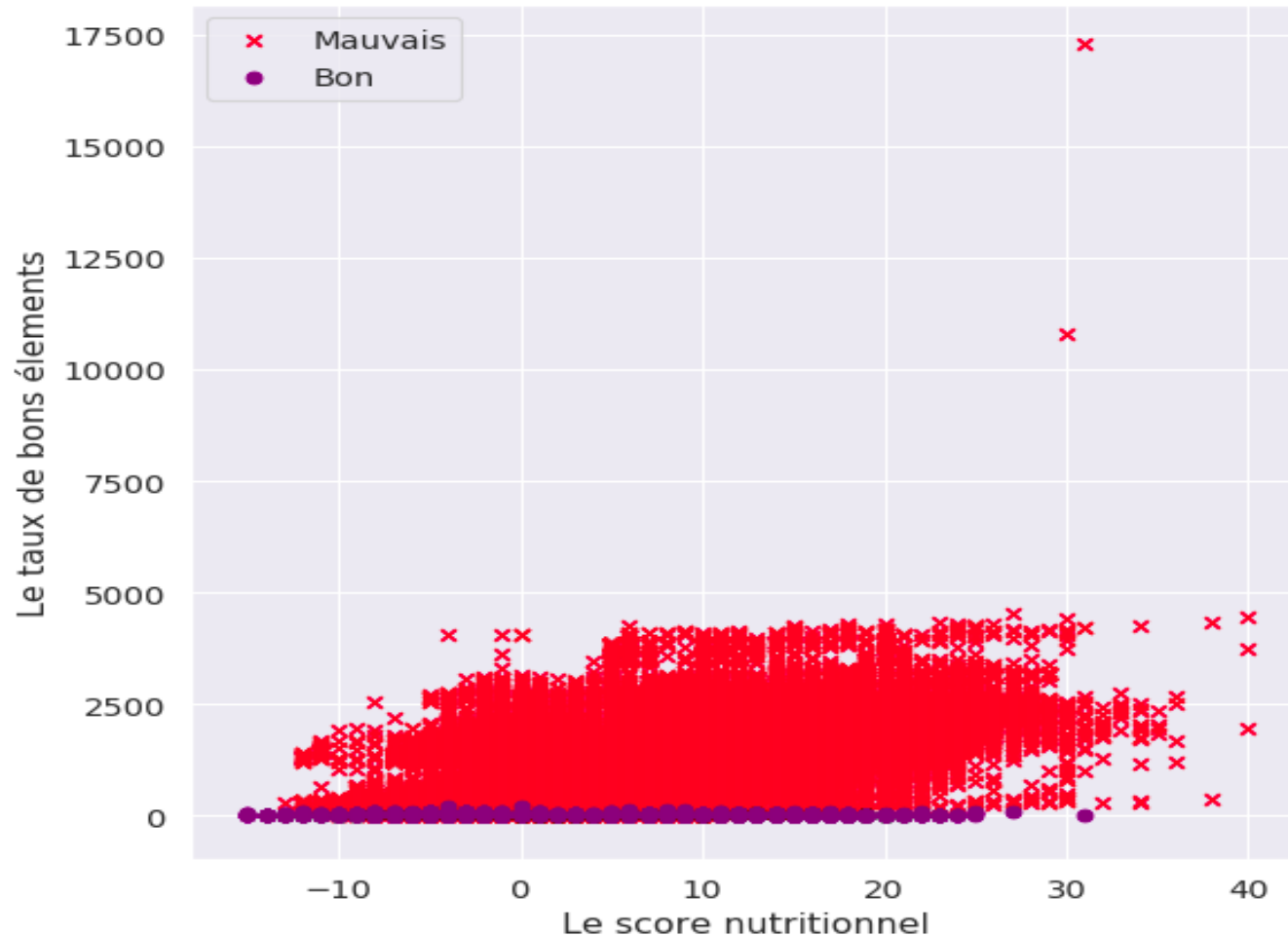
Features Engineering

- Deux features qui représentent le taux de valeur de la bonne nutrition et la mauvaise nutrition dans les produits.
 - Un feature qui donne une note à partir des deux features précédents aux produits.
- 

- Feature «bon_data_c» : taux global des fibres, protéines, calcium, vitamines, fer
- Feature «mauv_data_c» : taux globale des sucres, gras, cholestérol, énergie, sodium, gras saturés.
- Feature «score» : 'Bon' ou 'Mauvais'

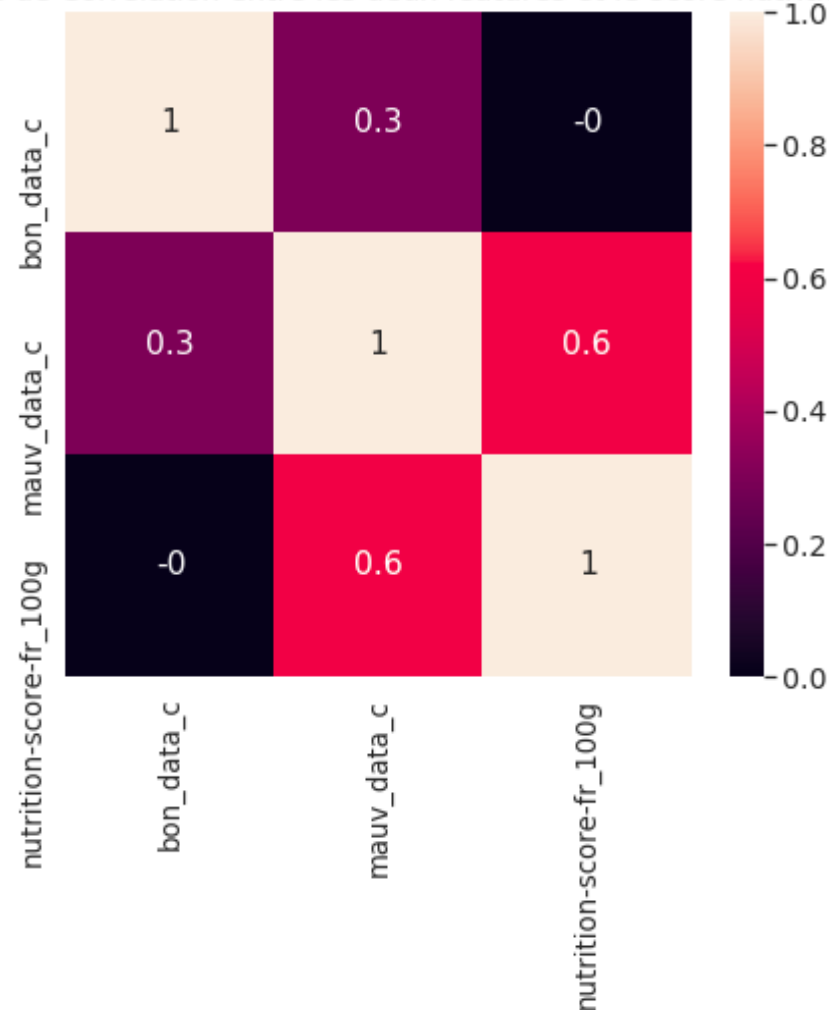
L'application des Features sur les données

La distribution des éléments selon leur score nutritionnel et leur nouveau feaature

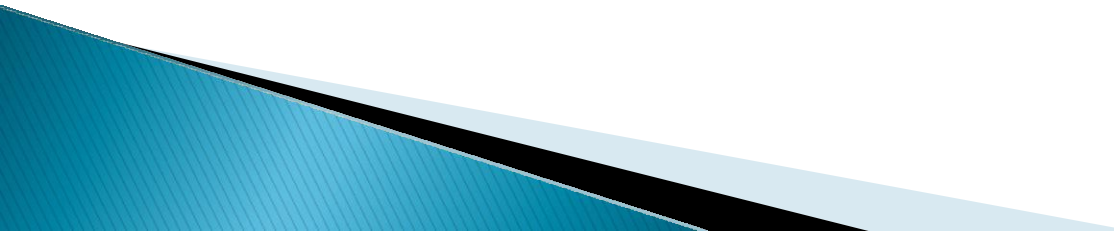


Corrélation avec le score nutritionnel

La matrice de Corrélation entre les deux features et le score nutritionnel



Réduction dimensionnelle

- Conversion des données vers un format standard
 - Réduire les dimensions en utilisant PCA
(7 composants)
 - 99.687% de la variance expliquée avec 7 composants
- 

Conclusion

- Une analyse de données exploratoire a permis de filtrer et interpréter les données importantes et voir la relation entre elles.
 - Création de features qui modélisent l'alimentation saine au sein des données.
- 